Technical Research Report 1171

TRAINING INDIVIDUAL IMAGE INTERPRETERS USING TEAM CONSENSUS FEEDBACK

John T. Cockrell
System Development Corporation

and

Robert Sadacca
Behavior and Systems Research Laboratory

U. S. Army
Behavior and Systems Research Laboratory

June 1971

BEHAVIOR AND SYSTEMS RESEARCH LABORATORY

An activity of the Chief, Research and Development

J. E. UHLANER
Director

Research accomplished
Under Contract to Department of the Army

System Development Corporation

NOTICES

DISTRIBUTION: Primary distribution of this report has been made by BESRL. Please address correspondence concerning distribution of reports to: U. S. Army Behavior and Systems Research Laboratory, Attn: RDMR-BLZ, 1300 Wilson Boulevard, Arlington, Virginia 22209.

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the Behavior and Systams Resaarch Laboratory.

NOTE: The findings in this report are not to be construed as an official Department of the Army position, unless so designated by othar authorized documents.

Technical Research Report 1171

AD

TRAINING INDIVIDUAL IMAGE INTERPRETERS USING TEAM CONSENSUS FEEDBACK

John T. Cockrell
System Development Corporation

and

Robert Sadacca
Behavior and Systems Research Laboratory

Abraham H. Birnbaum, Program Director

SUPPORT SYSTEMS RESEARCH DIVISION
Joseph Zeidner, Chief

BEHAVIOR AND SYSTEMS RESEARCH LABORATORY

Office, Chief of Research and Development
Department of the Army

1300 Wilson Boulevard, Arlington, Virginia 22209

June 1971

Army Project Number
2Q662704A721

Contract Number
DAHC19-67-C-0040
Surveillance Systems c-21

BESRL Technical Research Reports and Technical Research Notes are intended for sponsors of R&D tasks and other research and military agencies. Any findings ready for implementation at the time of publication are presented in the latter part of the Brief. Upon completion of a major phase of the task, formal recommendations for official action normally are conveyed to appropriate military agencies by briefing or Disposition Form.

FOREWORD

The SURVEILLANCE SYSTEMS research program of the U. S. Army Behavior and Systems Research Laboratory has as its objective the production of scientific data bearing on the extraction of information from surveillance displays, and the efficient storage, retrieval, and transmission of this information within an advanced computerized interpretation facility. Results are used in the development of techniques to enhance all phases of the image interpretation process and to provide findings for use in future systems design. Research is conducted under RDT&E Project 2Q662704A721, Surveillance Systems, FY 1971 Work Program.

The ADVANCED SURVEILLANCE SYSTEMS Work Unit continues BESRL research previously conducted as the COMPONENT INTEGRATION Work Unit, with emphasis on the human factor requirements of tactical interpretation facilities, as these facilities are currently evolving, particularly with regard to the critical analysis and evaluation of the functions of the interpreter, his display needs, and the effectiveness of the total system. The present publication summarizes four experiments leading to more effective methods of maintaining the skills of image interpreters within an advanced interpretation system through feedback furnished by team members.

BESRL research in this area is conducted as an in-house research effort augmented by contracts with organizations selected as having unique capabilities and facilities for research in aerial surveillance. The present research was conducted jointly by personnel of BESRL and of the System Development Corporation.

J. E. UHLANER, Director
Behavior and Systems
Research Laboratory

TRAINING INDIVIDUAL IMAGE INTERPRETERS USING TEAM CONSENSUS FEEDBACK

BRIEF

Requirement:

To explore the effectiveness of a low cost proficiency improvement and maintenance program for image interpreters; specifically, to determine the effect on interpreter detection and identification skills of a training method using feedback arising from the agreement (consensus) which team members achieve when they compare and discuss their interpretations.

Procedure:

Four experiments were conducted, exploring various team consensus feedback methods with varying conditions of team size and composition. In Experiment I, team consensus feedback was compared with precise feedback (prior identifications made by experienced interpreters) and no feedback in effectiveness in improving both target detection and target identification. Experiment II studied the effectiveness of team consensus feedback in improving target detection performance only. Experiment III evaluated modified team consensus feedback techniques designed to minimize feedback delays. In Experiment IV, the nature of learning curves using team consensus feedback techniques was investigated and the most promising procedures identified in the first three experiments were compared in a thorough analysis of the team consensus method.

Findings:

Team consensus feedback, requiring only that interpreters compare their answers in the detection and identification of tactical targets on previously interpreted or uninterpreted aerial imagery, was demonstrated to provide greater performance improvement than a control condition where interpreters practiced individually with no feedback. The greatest performance gains were made in target identification. The number of inventive errors was reduced using team consensus feedback training; improvement in the number of targets correctly detected was found only in the fourth experiment in which the most promising techniques of the first three experiments were compared. The results further showed that interpreters who were initially low in proficiency achieved the most significant gains in performance.

Interpreters assigned to teams that were heterogeneous in terms of initial proficiency made greater gains than interpreters that were members of teams homogeneous in terms of initial proficiency. Low proficiency interpreters were shown to make greater gains in proficiency than interpreters with high initial proficiency. The results indicate that low proficiency interpreters learned from, or benefited through their collaboration with, higher proficiency interpreters.

Team discussion and team size did not significantly influence interpreter performance; however, there was some indication that these variables combine to produce an overall effect.

The greatest performance gains were achieved with a method using three-man teams, heterogeneous in terms of initial proficiency, in which individuals made their initial interpretations individually but were permitted to discuss their identifications freely. Team methods requiring the continued close association of interpreters from the beginning of search did not lead to overall performance improvement, possibly because high proficiency interpreters precluded effort on the part of less proficient interpreters.

Utilization of Findings:

Team consensus feedback was demonstrated to be an effective method of improving and maintaining image interpreter proficiency in an operational image interpretation facility. The method requires no specialized training materials or instructor preparation. Two or three interpreters, using a roll of tactical imagery, may improve their performance, particularly in identification, by individually interpreting frames of imagery and then discussing their detections and identifications. More proficient interpreters should be included in each team to effect the greatest improvement in the less proficient interpreters. This method could also profitably be used in the classroom during team exercises.

CONTENTS

TABLES

FIGURES

TRAINING INDIVIDUAL IMAGE INTERPRETERS USING TEAM CONSENSUS FEEDBACK

BACKGROUND AND PURPOSE

Context of the Study

A series of experiments has been conducted over the past several years by the Behavior and Systems Research Laboratory (BESRL) to develop and test the team consensus feedback method as a technique for maintaining and enhancing the proficiency of individual image interpreters. The essential feature of the method is that interpreters practice in teams, arriving at decisions with regard to target detection and identification by a consensus of the team members. Generally, a consensus has been defined as either a two-thirds vote or a unanimous vote of the team members. The only feedback which the interpreters receive with regard to the accuracy of their decisions they provide for themselves during team discussion and comparison.

Image interpreters working alone on a mission are often unaware when they are doing a poor job of detecting or identifying targets. Seldom do they receive any feedback, and if they do, it is generally too late to be effective. In teams, however, interpreters can be forced continually to take stock of themselves, since their teammates are finding targets and making identifications which disagree with their own. In any conflict regarding identification of a specific target, someone has to be wrong. This awareness of disagreement not only forces team members to take a long hard look at the target, but also allows less proficient interpreters to become aware of some of their own deficiencies and to learn from the more proficient interpreters.

The team consensus feedback method is based on prior team studies conducted at BESRL [1, 2, 3], which demonstrated that image interpreters working in teams can produce more complete and accurate intelligence information from aerial reconnaissance imagery than interpreters working alone. The consensual judgment of team members was found to be especially effective in reducing the number of identification errors made by single interpreters. Since teams generally produce better reports than

---

[1] Sadacca, R., H. Martinek, and A. I. Schwartz. Image Interpretation Task--Status Report. Technical Research Report 1129, June 1962 (AD 411 228). Behavior and Systems Research Laboratory, Arlington, Va.

[2] Bolin, S. F., R. Sadacca, and H. Martinek. Team procedures in image interpretation. Technical Research Note 164, December 1965 (AD 480533). Behavior and Systems Research Laboratory, Arlington, Va.

[3] Doten, G. W., J. T. Cockrell, and R. Sadacca. The use of teams in image interpretation: Information exchange, confidence, and resolving disagreements. Technical Research Report 1151, October 1966 (AD 643312). Behavior and Systems Research Laboratory, Arlington, Va.

individuals, interpreters training in teams should receive more accurate knowledge of results, or feedback, than interpreters practicing alone.

The research approach followed has been to evaluate empirically various team consensus feedback procedures to determine which variations are most effective in training individual interpreters. Factors other than work procedures which have been varied in these experiments are delay of feedback, manner in which the feedback is presented, type of feedback, size of team, composition of the team in terms of initial proficiency of team members, and the effect of initial proficiency on subsequent learning. In the final experiment, reported here as Experiment IV, an attempt was made to investigate the nature and extent of the learning which occurs and the shape of the learning curve.

OBJECTIVES

Most field interpretation units typically have a relatively large number of inexperienced personnel and a relatively small number of experienced personnel. Some proficiency improvement and maintenance program is necessary for these personnel, especially for recent graduates of interpretation schools and transferees. The team consensus feedback method, if proved feasible, would offer a relatively simple and inexpensive method of providing this practice. The advantages of the method are that no elaborate and expensive materials are needed, and practice sessions can be initiated during any slack period by simply taking a roll of imagery "off the shelf" and presenting it to interpreter teams in an image interpretation facility.

The four experiments reported here were conducted in an effort to develop team practice methods which will lead to the greatest performance gains by individual interpreters. The first experiment was designed to obtain a general assessment of the usefulness of the consensus feedback process. In this experiment, team feedback was contrasted with no feedback on the one hand and precise feedback on the other. In the no-feedback condition, interpreters simply practiced by themselves. Precise feedback consisted of the "school solution" derived by a team of expert interpreters working in part with ground survey data. It was assumed that the quality of team consensus feedback would lie somewhere between no feedback and precise feedback, depending upon the effectiveness of the team procedure being used.

The second experiment sought to determine if team consensus feedback would be effective in enhancing detection performance. In the first experiment, most of the team practice had been concentrated on the identification of targets. In the second experiment, emphasis was shifted simply by eliminating the requirement for identification, thereby greatly increasing the amount of team practice in target detection.

In the third experiment, emphasis was concentrated once again on target detection, but team procedures were changed to minimize delay of feedback. This experiment was also structured to allow team output during

the practice sessions to be measured. This latter step was taken in order to determine if teams were detecting more targets during practice than were the interpreters working alone.

The fourth experiment was concerned mainly with the nature of the learning which occurs during the team consensus feedback practice sessions and also with the shape of the learning curve. Since this experiment was intended to be the final experiment in the series, it was considered a replication of some of the preceding experiments. To this end, the most effective procedures identified in the previous three experiments were evaluated.

FRAMEWORK FOR THE EXPERIMENTATION

The same general sequence of events was followed in all four experiments. Each group of subjects spent five days in laboratory experimentation. The first day was taken up with instructions and an individually administered pre-training test. The second, third, and fourth days were used for team practice sessions, and the fifth day was used for the post-training test. In the last experiment, a third individual test was introduced in the middle of the third day. For all experiments, a control group of individual interpreters practiced alone during the second, third, and fourth days on the same material used by the teams.

The imagery used in the experiments consisted of aerial photographs of Army field maneuvers. The imagery was arranged in stereo pairs, and 150 of these pairs were spliced together in a roll of 9" x 9" positive transparencies. Each interpreter had his own viewing device--a light table--and other basic interpreter equipment. The task consisted of searching each stereo pair for designated types of military targets, annotating the imagery by circling and numbering the targets, and then identifying the targets where identification was required. The interpreters then recorded this information on data sheets.

The experimental sample was composed of enlisted men who had just completed the image interpretation course at the U. S. Army Intelligence School, Fort Holabird, Maryland. These relatively inexperienced interpreters were judged to have proficiency consonant with the proficiency levels of interpreters who might benefit from consensus feedback exercises in the field. Fifty-four interpreters were used in the first experiment, 60 in the second, 48 in the third, and 54 in the final experiment.

FINDINGS

Effects of Different Types of Feedback (Experiment I)

In the first experiment, team consensus feedback was compared to precise feedback and no feedback. All teams in this experiment had three men assigned to each team, one man being high in proficiency. The remaining team members were either both medium in proficiency, both low in

proficiency, or one medium and one low. Proficiency was determined by means of the pre-training test. Three different consensus-feedback procedures were used in this experiment:

Serial Consensus. Each man initially worked independently on a different stereo pair. The men annotated all targets found, identified the targets, and recorded the targets on data forms. The team members then rotated seats and checked their teammates' work. Checking consisted of agreeing or disagreeing with the targets found by the previous interpreters and then looking for additional targets missed by the previous interpreters. Following checking, the three men got together and discussed all conflicts on each stereo pair.

Immediate Consensus. Each man worked on a copy of the same stereo pair during initial interpretation. The men then went directly to the discussion phase and attempted to resolve all conflicts.

Delayed Consensus. This method was the same as immediate consensus except that each man completed three stereo pairs before discussion was begun.

For the precise feedback condition, two work procedures were used:

1. Precise Team Procedure--The teams first went through the immediate consensus procedure and then were given the correct answers to compare with their team sheet.

2. Precise Individual Procedure--The men worked independently. Following interpretation of each stereo pair, each interpreter was given the correct answers to compare with his own.

For the no-feedback control condition, each man worked by himself and no feedback was received at any time.

In this first experiment, interpreters practicing under team consensus feedback made greater gains than interpreters in the no-feedback condition. As expected, those interpreters provided precise feedback showed the greatest gain. Another finding indicating the importance of feedback was that the delayed consensus procedure showed the least gain among the feedback procedures. This finding is in keeping with the general results in psychological experiments dealing with delay of reinforcement or feedback.

A breakdown of the results revealed that most of the gains in favor of the consensus feedback procedures occurred in identification of targets. No significant gains were obtained in the number of targets detected or in the number of false targets eliminated.

Effects of Consensus Feedback on Target Detection (Experiment II)

The second experiment explored the effect of consensus feedback on target detection. The immediate consensus procedure and the no feedback control procedure were used, but the requirement to identify targets was eliminated. This change reduced the amount of time required for each stereo pair and greatly increased the number of stereo pairs which could be covered during team practice, thereby presumably increasing the amount of detection practice. Another feature introduced in the second experiment was the use of confidence statements. Interpreters were told that if they recorded doubtful targets with a confidence of 49% or less, they would receive one-half credit if they were correct and would lose nothing if they were incorrect. This technique greatly encouraged responding during team practice and brought many more doubtful discriminations before the team for resolution. In addition, this experiment explored the effects of difference in size of team (2- and 3-man teams), the use of discussion versus no discussion, and team composition based on whether the team members were homogeneous or heterogeneous in initial ability.

Team consensus feedback greatly reduced the number of inventive 'false alarm' errors but did not significantly increase the number of targets correctly detected.

Results with regard to team size and discussion were non-significant, but results with regard to team type were significant. Interpreters who trained in heterogeneous teams showed greater gains than the homogeneous teams in both number of targets detected and reduction of inventive errors. Interpreters with initial low proficiency showed the greatest improvement in the reduction of inventive errors. Proficiency differences were significant for the reduction of inventive errors but not for the number of targets correctly detected.

Use of Joint Search Team Procedures to Minimize Delay in Consensus Feedback (Experiment III)

The third experiment explored two joint-search procedures designed to minimize the time interval between initial detection of a target and team consensus feedback. An equally important purpose of the joint-search procedures was to increase the number of targets detected by each team.

In the joint-search procedure with the least delay, each man had a separate light table and a copy of each stereo pair. As soon as one man detected a target, he would lean over and point to the target on the other man's light table. The men would then stop searching and discuss the target. After consensus had been reached, they would resume search until the next target was found. Generally, the first few easy targets were found and agreed upon in a very short time; increasing time intervals and longer discussions were noted as the targets became more doubtful.

In the second joint-search procedure, the men pointed out targets to each other as soon as they were found, but no discussion took place until both men had finished searching the stereo pair. At this time, all located targets were discussed.

In addition to these two work procedures, the serial search procedure and a no-feedback condition used in the first experiment were included.

All teams in the third experiment were composed of two men, one high in initial proficiency and one low. Confidence estimates and scoring were as in the second experiment. During the teamwork sessions, half the teams were required to identify the targets at a gross name level; the other half had only to detect the targets. During the individual pre- and post-training tests, only detection was required.

No significant improvement in performance was achieved either in number of targets correctly detected or in reduction of inventive errors. Thus, it would appear that the notion that more immediate feedback increases learning was not borne out by this experiment. However, practice decreased the number of inventive errors made by the initially less proficient interpreters.

Nature and Scope of Learning During Team Consensus Feedback (Experiment IV)

In the first three experiments, no attempt was made to determine directly if learning was occurring during the three days of consensus feedback training. Learning was inferred from differences between experimental and control groups. In the final experiment, an attempt was made to determine the nature and shape of the learning curve involved in team consensus feedback.

An effort was made to choose the team procedures and conditions which had led to the largest gains relative to the control group in the first three experiments. Two team procedures used in Experiment I, serial consensus and immediate consensus, were selected as the most promising of the procedures. In choosing team size, consideration was given to trends and other indications which seemed to indicate that three-man teams are preferable to two-men teams. Although no beneficial effects had been demonstrated for discussion, team members were allowed to discuss their target identifications. The teams were required to identify all targets at gross name level. All teams were heterogeneous in terms of initial proficiency, each team consisting of one high, one medium, and one low proficiency interpreter.

Three individual tests were used to assess learning: pre-training, intermediate, and post-training. The three tests were roughly equated as to difficulty, but to insure that difficulty did not bear on the results, a Latin square arrangement was used to counterbalance tests, sessions, and training procedures.

Results of the final experiment showed very definitely that learning occurred during the three-day practice period and that this learning was significantly greater for the team consensus feedback groups. Learning was demonstrated for all performance measures--target identification, number of targets correctly detected, and number of inventive errors.

The immediate consensus feedback procedure seemed slightly superior to the serial consensus procedure. With serial consensus feedback procedure, learning was rapid at first and then slowed sharply; learning with the immediate consensus feedback procedure was steady at a medium rate. For the control group, learning occurred at a low steady rate.

IMPLICATION OF THE FINDINGS

The general conclusion from the four experiments is that the team consensus feedback is an effective method for maintaining and improving image interpreter proficiency. Although precise feedback leads to greater learning, precise feedback is seldom available, especially in field situations. Consensus feedback training can be readily implemented where two or more interpreters have free time and a roll of imagery.

Overall results for the four experiments indicate that team consensus feedback practice leads to greater learning than individual practice in target detection and identification. The method has been shown to be most effective in increasing identification proficiency. The team consensus feedback method was also effective in reducing number of false targets reported. Only with the most effective team procedures (in Experiment IV) was there a significant increase in number of targets correctly detected.

Team composition in terms of proficiency seems highly important in the team consensus feedback procedure. Where teams are heterogeneous in proficiency, learning is substantially greater than where teams are homogeneous in proficiency. This fact seems to indicate that team members learn from each other, and unless at least one member of the team is more proficient than the others, very little learning will occur. This conclusion was clearly borne out in Experiment IV where the highly proficient interpreters in the team practice groups learned only a small amount more than the control interpreters, whereas the interpreters of medium and low proficiency learned much more than the control interpreters.

The ideal procedure seems to be to have three-man teams whose members are heterogeneous in proficiency using the immediate consensus work procedure. In this procedure, all interpreters first perform initial interpretation on an individual basis and then get together to decide on the team report. The crucial factors seem to be the individualized initial interpretation and the opportunity for immediate feedback through comparison of responses and discussion with teammates.

TECHNICAL SUPPLEMENT

METHOD

Orientation and Instructions

In all four experiments, the first half-day was taken up with orientation and instructions. Following the first presentation of the instructions, the interpreters were given practice frames containing targets that could easily be identified. Each of the practice data sheets was checked individually to determine if the interpreters understood the instructions. The instructions were then repeated for any procedure which had been misunderstood by one of the interpreters. During this period and subsequently throughout the experiments, each interpreter was given a set of photographic keys which contained photographs, scale drawings, and measurements for each target on the target list. The photographic keys also contained vertical photographs of each target in stereo at a scale similar to those used in the experiments. To insure that subjects were thoroughly familiar with the keys, an orientation test consisting of several frames containing annotated targets was given the subjects. These orientation tests contained one or more targets for each type of target on the target list. To identify these targets, it was necessary to study the keys thoroughly. Following practice with the keys, subjects were given another orientation test consisting of several frames of imagery similar to that used during the main experiment to familiarize them with the target detection problems they would encounter during the pre-training test. During the instruction period, no feedback of any kind was given with reference to any of the targets on the imagery.

Pre-training Test

This test was given individually to all interpreters in all experiments on the second half of the first day. Generally, the test consisted of 9-12 stereo pairs containing a total of 40-75 targets. The stereo pairs were selected arbitrarily from the pool of 150 stereo pairs. An effort was made to have all scales represented and to have frames with varying numbers of targets. The tests were different for each experiment although there was some overlap. Ten minutes were allowed for completion of each stereo pair, and interpreters were started and stopped in unison. In the event an interpreter finished prior to the end of the 10-minute period, he was required to sit quietly until the next stereo pair was started. No feedback was given during the test, and interpreters were not permitted to talk about the test during breaks nor to compare data sheets.

Team Practice

Team practice was conducted during the second, third, and fourth experimental days. The control group practiced as individuals on copies of the imagery used by the teams. In Experiment I, all teams were required to start and stop each set of three stereo pairs at the same time to control the amount of material covered by each team. During Experiments II, III, and IV, the teams were allowed to proceed at their own pace to control the total amount of time spent in practice.

Some of the work procedures used by the teams were common to two or more experiments and some were unique. The procedures were:

1. Serial Consensus--This procedure was used in Experiments I, III, and IV. Each man performs initial interpretation on a different stereo pair. The men then rotate chairs, correcting and adding targets to their teammates' work. The men then discuss all conflicts.

2. Immediate Consensus--This procedure was used in Experiments I, II, and IV. Each man performs initial interpretation on a copy of the same stereo pair. The men then discuss all conflicts.

3. Delayed Consensus--This procedure was only used in Experiment I and is the same as immediate consensus except that team members individually perform initial interpretation on three stereo pairs and then discuss all three at once.

4. Joint-Search--This procedure was only used in Experiment III. Team members perform initial interpretation together, pointing out targets to each other as they are found. In one variation, the teammates discuss each target as it is found; in the other, the teammates wait until they have completed the stereo pair to discuss the targets.

5. Control Procedure--Each man performs the interpretation by himself and does not discuss his targets or compare response sheets with other interpreters. The control procedure was the same in all experiments.

Post-training and Intermediate Tests

A post-training test was given in all experiments. This test was given on the fifth day and generally lasted most of the day. The test consisted of 9-17 stereo pairs containing a total of 40-100 targets. The nature of the post-training test and the manner of administration were similar to the pre-training test. In Experiment IV, an intermediate test was given during the middle of the third day.

Experimental Subjects

Image interpreter trainees just graduated from the image interpretation course at the U. S. Army Intelligence School, Fort Holabird, Maryland, were the subjects for the four experiments.

Experimental Imagery

The imagery used in the experiments were aerial photographs of Army field maneuvers. The photos were arranged in stereo pairs; 150 of these pairs were spliced together in a 9" x 9" roll of positive transparencies. There was a stereo overlap of 40-60% on each stereo pair, and scales ranged from 1:1000-1:5000. Each stereo pair contained from 0-19 targets.

Confidence Statement

The interpreters in all four experiments were required to place a confidence estimate ranging from 0-100% after each target. In Experiment I, confidence did not enter into the scoring, but in Experiments II, III, and IV the interpreters were told that confidence would affect their scores. Confidence was only used with reference to target detection and did not pertain to identification. For Experiments II and III, confidence estimates ranging from 50-100% were scored with a weight of +1 for real targets and -1 for false targets. Confidence estimates ranging from 0-49% were scored with a weight of +1/2 for real targets and 0 for false targets. In Experiment IV, an attempt was made to expand the weights and use the full scale of confidence as follows:

Confidence	Real Targets	False Targets
100	+4	-4
90	+4	-3
80	+4	-3
70	+4	-2
60	+3	-2
50	+3	-1
40	+2	-1
30	+2	-1
20	+1	-0
10	+1	-0
0	+0	-0

Dependent Variables

The three basic scores used were based on number of correct identifications, number of targets correctly detected, and number of inventive errors (false targets). All scores were weighted, the weights being based either on target priority (Experiment I) or confidence estimates

(Experiments II, III, and IV). In Experiment I, offensive tracked vehicles and artillery were given greater weight than other targets. (This weighting scheme had little absolute effect and no differential effect on the results, and the scheme was dropped in later experiments.) In Experiments II, III, and IV, weights were based on the confidence estimate for each target reported. This weighting scheme also had little differential effect on the scores; however, it was retained in order to encourage as much responding as possible.

The identification score was computed by adding up points for each target correctly identified and subtracting a point for each target incorrectly identified. Targets detected but not identified did not enter into the identification score. The correct detection score was computed by summing points for each target on the target list correctly detected. The inventive error score was computed by summing the points for all false targets reported. In Experiment III, these last two scores were converted to percentage scores as shown below to allow comparison between individual scores and team score:

$$\text{Detection completeness} = \frac{\text{Sum of points for targets correctly reported}}{\text{Sum of points for all possible targets}} \times 100$$

$$\text{Detection accuracy} = \frac{\text{Sum of points for targets correctly reported}}{\text{Sum of points for targets correctly reported} + \text{points for false targets reported}} \times 100$$

EXPERIMENT I. EFFECTS OF DIFFERENT TYPES OF FEEDBACK[4]

The primary objective of the first experiment was to determine the effectiveness of team consensus feedback for improving interpreter proficiency. Team consensus feedback was compared with precise feedback and no feedback. The experiment involved six experimental conditions:

Team Consensus Feedback Variations

1. Serial consensus feedback
2. Immediate consensus feedback
3. Delayed consensus feedback

---

[4] A more complete description of this experiment may be found in Cockrell, J. T. Maintaining image interpreter proficiency through team consensus feedback. Technical Research Note 195. April 1968 (AD 833583). Behavior and Systems Research Laboratory. Arlington, VA.

Precise Feedback Variations

4. Precise team feedback
5. Precise individual feedback

Control Group

6. No feedback

Experimental Design

The six experimental groups were compared by means of difference scores measuring gain in individual proficiency over pre-training performance. Nine interpreters were assigned to each method. The 54 interpreters were categorized as to proficiency on the basis of their initial performance scores. A randomized-blocks analysis of variance was used to assess the significance of the differences between initial and final performance scores across methods and proficiency groups.

The nine interpreters assigned to each of the four team methods were formed into three-man teams of three types: high-medium-medium, high-medium-low, and high-low-low. Assignment of men at each proficiency level to a team was on a random basis. One high proficiency man was assigned to each team type on the hypothesis that less proficient interpreters would benefit from working with more proficient interpreters in the consensus feedback methods.

Results for Experiment I

The main results showing the effect of the three different types of feedback on interpreter proficiency are presented in Table 1. The scores are generally in the predicted direction; namely, the precise feedback methods showed the greatest overall gain, the consensus feedback method showed intermediate gain, and the no-feedback method showed the least gain. Tables 2, 3, and 4 show the three sets of detection and identification scores arranged according to proficiency level. Table 5 shows the F-ratios for each of the scores.

Practice involving feedback improved only target identification significantly and not detection. Further analysis of the identification scores (using Dunnett's test) indicated that performance under the precise team feedback condition was significantly better than for the no-feedback condition at the .01 level and that performance under the precise individual feedback and serial consensus feedback conditions was better than that under the no-feedback condition at the .05 level. Differences among the various team consensus and precise feedback conditions were not significant.

Table 1

MEAN DIFFERENCES BETWEEN PRE-TEST AND POST-TEST SCORES FOR THREE MEASURES OF IMAGE INTERPRETER PERFORMANCE--EXPERIMENT I

Method	Scores		
	Identification	Correct Detection	False Target Detection[a]
Precise Feedback	51.2	73.9	21.6
Consensus Feedback	41.1	71.9	19.1
Control	28.0	69.4	15.1

[a] Positive value represents a decrease in number of false target responses.

EXPERIMENT II. EFFECTS OF CONSENSUS FEEDBACK ON TARGET DETECTION [5]

The primary objective of the second experiment was to determine the effect of team consensus feedback on target detection. In Experiment I, there was no significant difference between the control group and experimental groups on the detection skill. It was felt that this may have been due to the lack of sufficient practice on this skill. Accordingly, in Experiment II, the requirement for target identification was eliminated, and interpreters were required only to indicate the location of each target and their confidence that it was truly a target. Each team in Experiment II completed approximately four times as many practice stereo pairs as the teams in Experiment I and thereby presumably increased amount of detection practice.

[5] A more complete description of this experiment may be found in Cockrell, J. T. Maintaining Target Detection Proficiency Through Team Consensus Feedback. Technical Research Note 219. October 1969 (AD 707376). Behavior and Systems Research Laboratory. Arlington, VA.

Table 2

MEAN PERFORMANCE GAINS IN IDENTIFICATION SCORE BETWEEN
PRE-TEST AND POST-TEST--EXPERIMENT I

Proficiency Level	Training Method: Precise Feedback		Consensus Feedback			Control	All Methods
	Team	Individual	Serial	Immediate	Delayed	No Feedback	
High	37.7	56.0	57.7	49.0	36.7	33.0	45.0
Medium	57.7	62.3	48.3	24.3	34.3	20.3	41.2
Low	67.3	26.3	44.7	34.3	40.7	30.7	40.7
All Levels	54.2	48.2	50.2	35.9	37.2	28.0	

Table 3

MEAN PERFORMANCE GAINS IN CORRECT DETECTION SCORE BETWEEN
PRE-TEST AND POST-TEST--EXPERIMENT I

Proficiency Level	Training Method: Precise Feedback		Consensus Feedback			Control	All Methods
	Team	Individual	Serial	Immediate	Delayed	No Feedback	
High	70.0	81.0	89.7	88.3	62.3	66.0	75.1
Medium	65.0	80.0	83.7	77.0	68.0	64.3	73.0
Low	74.0	73.3	51.3	66.3	61.0	84.7	68.4
All Levels	69.7	78.1	74.9	77.2	63.8	69.4	

Table 4

MEAN PERFORMANCE GAINS IN FALSE TARGET DETECTION SCORE[a] BETWEEN PRE-TEST AND POST-TEST--EXPERIMENT I

Proficiency Level	Training Method: Precise Feedback		Consensus Feedback			Control	All
	Team	Individual	Serial	Immediate	Delayed	No Feedback	Methods
High	16.3	10.7	18.0	6.3	6.7	12.0	11.3
Medium	26.7	18.7	34.7	23.0	12.7	19.3	22.5
Low	24.7	32.7	35.7	23.0	13.7	14.0	23.9
All Levels	22.5	20.7	28.8	17.4	11.0	15.1	

[a] Positive score indicates decrease in the number of false target responses.

Table 5

F-RATIOS FOR DIFFERENCE SCORES--EXPERIMENT I

Source	df	Score: Identification	Score: Correct Detection	Score: False Target Detection
Methods (M)	5	2.64*	1.33	1.25
Proficiency (P)	2	.29	.86	.31
MP	10	1.41	1.75	.80
Within (Mean Square)	36	349.1	242.1	276.7
Total	53			

*P < .05.

Experimental Design

In Experiment II, in addition to a detailed study of the effects of team consensus feedback on target detection, four other factors were studied: 1) team size (2-man or 3-man),2) team composition (team members homogeneous in initial ability versus heterogeneous), 3) discussion versus no discussion, and 4) interpreter proficiency (high, medium, and low). The design used for this experiment was a 2x2x2x3 factorial with a control group. The detailed breakdown is shown in Table 6.

Performance among the eight experimental groups and the no-feedback (control) group, shown in the columns in Table 6, was compared by means of difference scores obtained by subtracting pre-training test scores from post-training test scores. A randomized-blocks analysis of variance was used to assess the significance of differences among the experimental conditions. The three blocks were based on initial proficiency as measured by the pre-training test. From each block, two men were drawn randomly for each of the experimental groups. The six interpreters assigned to each of the eight experimental groups were assigned to homogeneous or heterogeneous teams and to two-man or three-man teams as shown in Table 7. The control group included 12 interpreters.

Table 6

EXPERIMENTAL DESIGN FOR EXPERIMENT II
(N = 2 Interpreters in Each Cell, 12 Control Subjects, 48 Total)

Proficiency Group	3-Man Teams				2-Man Teams				Total	Control
	Discussion		No Discussion		Discussion		No Discussion			
	HOM	HET	HOM	HET	HOM	HET	HOM	HET		
High	2	2	2	2	2	2	2	2	16	4
Medium	2	2	2	2	2	2	2	2	16	4
Low	2	2	2	2	2	2	2	2	16	4
Total	6	6	6	6	6	6	6	6	48	12

Table 7

COMPOSITION BY PROFICIENCY OF TEAMS IN EXPERIMENT II

Team Type	3-Man Teams	2-Man Teams
Homogeneous	1. High, High, Medium	1. High, High
	2. Medium, Low, Low	2. Medium, Medium
		3. Low, Low
Heterogeneous	1. High, Medium, Low	1. High, Low
	2. High, Medium, Low	2. High, Medium
		3. Medium, Low

Team Procedures

Team procedures used in Experiment II were two variations on the immediate consensus procedure. Where discussion was used, the procedure was the same as in Experiment I. Where discussion was not used, the teammates merely passed their completed data forms to each other so that comparisons could be made.

Results for Experiment II

The results for Experiment II were analyzed in two steps. The first step compared the differences among the various experimental factors; the second step compared the team consensus feedback groups to the control group. Table 8 shows the mean performance gains for the various factors in terms of scores based on number of correct detections and number of inventive errors. This table shows a substantial difference in favor of heterogeneous teams for both scores, and a large reduction in the number of inventive errors by low and medium proficiency interpreters.

Table 9 presents the analysis of variance results for the four factors considered. Team size and discussion versus no-discussion comparisons showed no significant performance increases. Interpreter proficiency (high, medium, and low) showed a significant difference in the reduction of inventive errors; low proficiency interpreters showed a marked reduction in inventive errors, while more proficient interpreters

showed less marked improvement. The performance of heterogeneous teams improved significantly more than that of homogeneous teams both in number of targets correctly detected and number of inventive errors. From these last two sets of results, there is the indication that low proficiency interpreters gain from the team consensus feedback method and that they benefit from their association with a more proficient team member.

In order to analyze the difference between the no-feedback control group and the consensus feedback experimental groups, the discussion and team-size factors were collapsed, leaving only team type and proficiency factors. Table 10 shows the mean difference scores for this analysis in both detection scores. Table 11 shows the analysis of variance for groups with unequal numbers of subjects. Team consensus feedback training resulted in a significant reduction in the number of inventive errors, when compared with performance of the control group. Team consensus feedback, however, did not improve detection performance in terms of number of targets detected. The results in Table 11 also repeat the finding of significant gains in error reduction for the low proficiency interpreters.

EXPERIMENT III. JOINT SEARCH PROCEDURES TO MINIMIZE DELAY IN CONSENSUS FEEDBACK

In Experiment III, two new teamwork procedures were designed to minimize the time delay between initial interpretation and consensus feedback. These procedures used joint target search rather than individual search. A reduction in feedback time delay was felt to be important in view of the results from Experiment I with delayed consensus feedback.

Team Methods

The two-man team procedures were as follows:

Joint Search -- One-Response Delay. Interpreters search together and discuss each target as it is detected.

Joint Search -- One-Frame Delay. Interpreters search together but discuss all targets at end of frame.

Serial Search -- Two-Frame Delay. Interpreters search individually on different frames, cross-check by swapping seats, then discuss each frame in turn.

Independent Search -- No Feedback. In this control group, interpreters search individually with no feedback at any time.

During team practice sessions, one half the teams were required to make a gross identification of all targets (e.g., tracked vehicle, truck, shelter). This gross identification was not required during the pre-training and post-training tests. The purpose of the identification

Table 8

MEAN GAINS BETWEEN PRE-TEST AND POST-TEST
IN DETECTION SCORES--EXPERIMENT II

Experimental Factors	Correct Detection	False Target Detection
Team Type		
Heterogeneous	66.4*	9.7*
Homogeneous	56.3	0.9
Discussion		
Yes	61.0	4.1
No	61.8	6.6
Team Size		
3-Man	62.8	7.9
2-Man	60.0	2.8
Initial Proficiency		
High	66.4	0.2**
Medium	60.9	9.2
Low	56.9	19.8**

*$P < .05$.
**$P < .01$.

Table 9

DIFFERENCE SCORE F-RATIOS FOR DETECTION SCORES--EXPERIMENT II

Source	df	F-Ratios: Correct Detection	F-Ratios: False Target Detection
Team Type (T)	1	4.34*	6.89*
Proficiency (P)	2	1.28	12.37**
Discussion (D)	1	.02	.55
Team Size (S)	1	.31	2.36
TP	2	.68	.17
TD	1	.41	.63
TS	1	.61	.80
PD	2	.42	.45
PS	2	.08	3.26
DS	1	.56	.01
TPD	1	.05	.35
TPS	2	.01	1.08
TDS	1	.01	.55
PDS	2	3.39	4.23*
TPDS	2	.40	.12
Within (Mean Square)	24	283.4	135.8
Total	47		

*P < .05.

**P < .01.

Table 10

COMPARISONS BETWEEN EXPERIMENTAL GROUPS AND CONTROL GROUPS FOR MEAN GAINS IN DETECTION SCORES--EXPERIMENT II

Proficiency Group	Correct Detection		
	Heterogeneous Teams	Homogeneous Teams	Control
High	67.8	65.0	71.7
Medium	66.5	55.4	70.0
Low	65.1	48.6	65.5
All Groups	66.5	56.3	69.1

	False Target Detection[a]		
	Heterogeneous Teams	Homogeneous Teams	Control
High	0.2	-8.0	-29.0
Medium	9.2	-2.2	-23.0
Low	19.8	13.0	5.0
All Groups	9.8	-0.9	-19.2

[a]Positive score represents decreese in the number of felse target responses.

Table 11

F-RATIOS FOR COMPARISON BETWEEN COLLAPSED EXPERIMENTAL GROUPS AND CONTROL GROUPS IN DETECTION SCORES--EXPERIMENT II

Source	df	F-Ratios	
		Correct Detection	False Target Detection
Method (M)	2	3.01	12.12**
Proficiency (P)	2	1.18	6.73*
M x P	4	.30	.07
Within (Mean Square)	51	270.6	322.2
Total	59		

*P < .05.

**P < .01.

requirement was to determine if team output would be affected in terms of amount and accuracy. A count of the number of stereo pairs completed during the teamwork sessions showed that the identification requirement had no effect on amount of interpretation accomplished.

Experimental Design

For Experiment III, the variables studied were team method, requirement for identification (required versus not required), and initial proficiency (high versus low). The experimental design is a 4x2x2 factorial with pre-post test difference scores as dependent variable. Table 12 shows a detailed breakdown.

Results for Experiment III

While it could be expected from the rationale for the team consensus feedback method that variations providing more immediate feedback--such as the joint-search methods--would result in improved performance, this notion was not borne out by the results. Tables 13 and 14 show the pre-test, post-test, and difference percentage scores based on number of correct detections and number of inventive errors, respectively. Difference scores between initial and final proficiency--in terms of mean percentage gain--were used for the analysis of means, the results of which are shown in Table 15. There were no significant differences for feedback method, i.e., no demonstrated superiority of the more immediate feedback methods over other feedback methods. The only significant difference found was for initial proficiency; low proficiency interpreters showed significant reduction in the number of inventive errors made.

EXPERIMENT IV. NATURE OF LEARNING DURING TEAM CONSENSUS FEEDBACK

The purpose of Experiment IV was to determine the nature of the learning which occurs during the teamwork sessions, and to some extent, the shape of the learning curve. This experiment differed from the preceding experiments in that it permitted measurement of amount of learning independent of test difficulty. Three proficiency tests were used which were counterbalanced to eliminate the effects of test difficulty.

The procedures and team composition which had been found to be best in previous experiments were used -- three-man teams, heterogeneous teams in terms of initial ability, and discussion as team members wished. The work procedures were serial consensus and immediate consensus plus a control group.

Table 12

ARRANGEMENT FOR ANALYSIS OF VARIANCE OF DIFFERENCE SCORES BETWEEN INITIAL AND FINAL PROFICIENCY TESTS--EXPERIMENT III
(N = 3 Interpreters in Each Cell, 48 Total)

Proficiency Group	Joint Search				Serial Search		Independent Search (Control Group)		Total
	One-Response Delay		One-Frame Delay		Two-Frame Delay		No Feedback		
	ID	No ID	ID	No ID	ID	No ID	ID	No ID	
High	3	3	3	3	3	3	3	3	24
Low	3	3	3	3	3	3	3	3	24
Total	6	6	6	6	6	6	6	6	48

Table 13

MEAN PERCENTAGE SCORES FOR DETECTION COMPLETENESS
ON PRE-TEST AND POST-TEST--EXPERIMENT III

Factors	Proficiency Tests		
	Initial Test	Final Test	Difference
Team Method			
Joint-Search-One Response	42.9	51.0	8.1
Joint-Search-One Pair	42.5	54.3	11.8
Serial Search-Two Pairs	43.7	52.7	9.0
Independent Search	42.6	53.4	10.8
Identification			
Yes	43.1	54.4	11.3
No	42.8	51.3	8.5
Initial Proficiency			
High	48.5	57.6	9.1
Low	37.3	48.1	10.8

Table 14

MEAN PERCENTAGE SCORES FOR DETECTION ACCURACY ON PRE-TEST AND POST-TEST--EXPERIMENT III

Factors	Proficiency Tests		
	Initial Test	Final Test	Difference
Team Method			
Joint-Search-One Response	83.4	91.4	8.0
Joint-Search-One Pair	83.5	86.0	2.5
Serial Search	81.8	81.3	-0.5
Independent Search	81.7	83.9	2.2
Identification			
Yes	82.8	85.2	2.4
No	82.4	86.1	3.7
Initial Proficiency			
High	92.6	87.4	-5.2
Low	72.6	84.0	11.4

Table 15

F-RATIOS FOR DIFFERENCE SCORES BETWEEN PRE-TESTS AND POST-TEST--EXPERIMENT III

Source	df	Detection Completeness	Detection Accuracy
Team Method (M)	3	.77	.63
Identification (I)	1	2.02	.09
Initial Proficiency (P)	1	.69	13.39**
MI	3	2.84	.10
MP	3	1.19	1.32
IP	1	.40	1.63
MIP	3	.44	.53
Within (Mean Square)	32	49.93	246.15
Total	47		

**P < .01.

Experimental Design

Each subject was given one of the three individual performance tests as a pre-training test. Based on scores on this test, interpreters were divided into three proficiency blocks -- high, medium, and low. One man was randomly drawn from each block to form three-man teams and these teams were assigned randomly to one of the three work procedures. Table 16 shows the design for the experiment.

Results

Results for Experiment IV are summarized in Table 17. In general, improved performance was found across the three sessions, including that for the no-feedback (control) group. However, the two consensus feedback groups showed greater learning than the control group on all three performance measures -- identification, correct target detection, and inventive errors. Thus, the two consensus feedback methods were effective in improving interpreter proficiency. Figures 1, 2, and 3 present a breakdown of results by proficiency level. For all three performance variables, the greatest performance gains were achieved by low proficiency interpreters. Medium proficiency interpreters gained less, and high proficiency interpreters gained little more than interpreters in the control (no feedback) condition.

Table 18 shows the results of the analysis of variance for factors of primary experimental concern -- feedback method, initial proficiency, and sessions -- as well as for order and test. The analysis of variance was carried out on performance score (not performance gain as in prior experiments) for identification, correct detections, and inventive errors. The results presented in Table 18 show clearly that (initial) proficiency is an important factor, the low proficiency interpreters benefiting most from team consensus feedback training. The results also show that learning occurred, as evidenced by the significant sessions effect.

The apparent lack of significant results for method (except for the correct detection score) requires a word of explanation. As subjects were assigned to experimental methods on the basis of initial test scores, differences among methods across the three tests (pre-test, intermediate test, and post-test) would be reduced. For this reason, additional analyses were carried out to look in more detail at differences among feedback methods. The Newman-Keuls test (6) was used to compare the two feedback methods against the control group and against each other for each of the three tests separately as shown in Table 19. The application of the Newman-Keuls test to scores on the pre-test shows that, as desired, there were no significant differences among the the two feedback conditions and

---

[6] Winer, B. J. Statistical Principles in Experimental Design, McGraw-Hill Book Company, New York, 1962.

Table 16

EXPERIMENTAL ARRANGEMENT FOR EXPERIMENT IV
(N = 54 Subjects, 2 Per Cell)

Method	Order	Proficiency	Sessions		
			1	2	3
Immediate Consensus	I	H M L	Test A	Test B	Test C
	II	H M L	B	C	A
	III	H M L	C	A	B
Serial Consensus	I	H M L	A	B	C
	II	H M L	B	C	A
	III	H M L	C	A	B
Control	I	H M L	A	B	C
	II	H M L	B	C	A
	III	H M L	C	A	B

the control condition and between the feedback conditions, since feedback was not introduced in the pre-test condition. However, Table 19 does show significant differences between both feedback groups and the control group in both identification and correct detections in subsequent sessions. No significant differences were found for inventive errors. These results, then, again show that team consensus feedback is effective in achieving increased interpretation proficiency, particularly in target identification.

Table 17

MEAN SCORES FOR IDENTIFICATION, CORRECT DETECTION, AND INVENTIVE ERRORS ACROSS TEST ADMINISTRATIONS--EXPERIMENT IV

Dependent Variable	Method	Test Administration 1	2	3
Identification	Immediate Consensus	52.1	68.8	76.9
	Serial Consensus	49.8	71.5	74.7
	Control	50.6	59.9	66.7
	Average	50.8	66.7	72.8
Correct Detection	Immediate Consensus	86.6	102.3	117.6
	Serial Consensus	81.7	115.5	116.9
	Control	83.1	93.2	101.5
	Average	83.8	103.7	112.0
Inventive Errors	Immediate Consensus	-16.3	-11.0	-9.1
	Serial Consensus	-15.8	-16.8	-13.5
	Control	-13.9	-16.0	-14.9
	Average	-15.3	-14.6	-12.5

Table 18

ANALYSIS OF VARIANCE OF IDENTIFICATION, POSITIVE DETECTION, AND FALSE TARGET DETECTION--EXPERIMENT IV

Source of Variance	df	F-Ratios Identification	Correct Detection	False Target Detection
Between Subjects				
Method (M)	2	3.25	3.41*	1.79
Order (O)	2	12.17**	5.90**	5.23*
Proficiency (P)	3	17.36**	15.83*	4.05*
MO	4	1.04	1.90	2.50
MP	4	.81	.78	.98
OP	4	1.42	.48	.76
MOP	4	.72	.96	.71
Subjects Within Groups (Mean Sq.)	27	239.08	648.41	93.07
Within Subjects				
Sessions (S)	2	113.70**	60.00	3.48*
Test (T)	2	.41	1.88	.44
MS	4	3.50*	4.39**	3.07*
MT	4	.44	1.94	4.68**
PS	4	6.77	2.35	9.94**
PT	4	.67	.86	1.46
MPS	8	.46	1.36	.86
MPT	8	1.45	1.04	.49
Residual	18	2.24*	.64	1.58
Error (Mean Sq.)	54	61.12	189.03	34.12
Total	108			

*P < .05.

**P < .01.

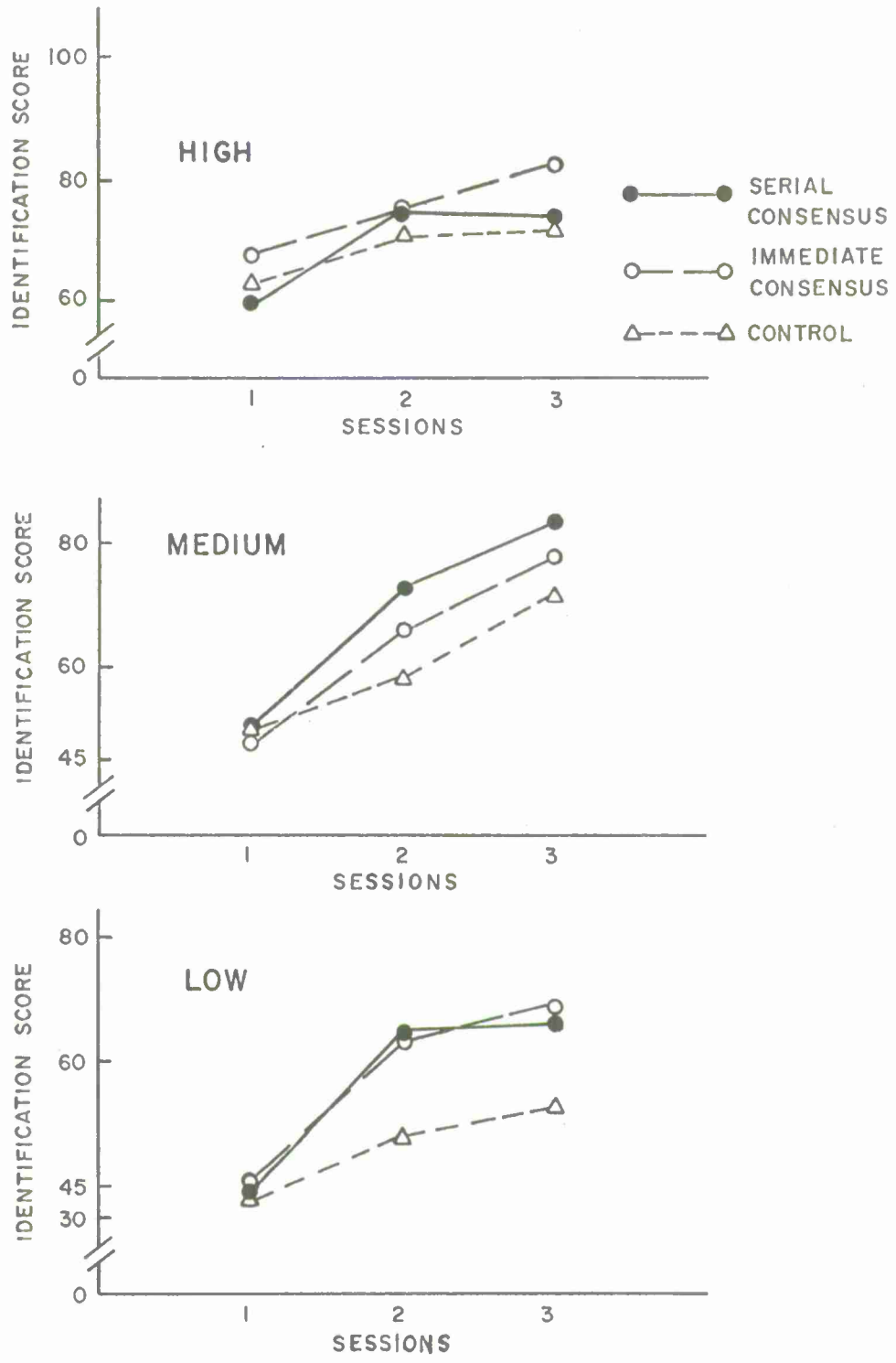

Figure 1. Identification learning curves for three methods used by interpreters of high, medium, and low proficiency.

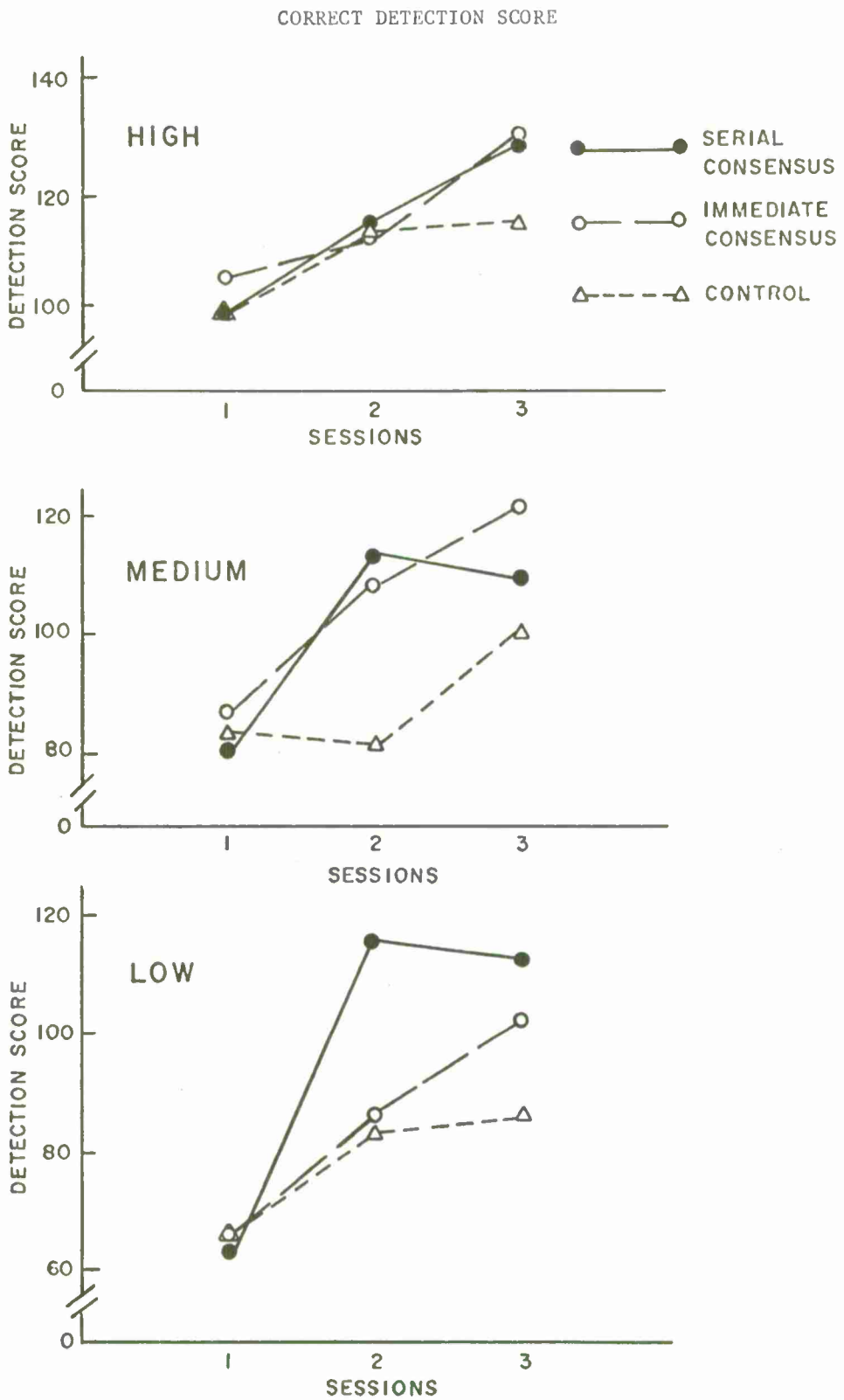

Figure 2. Detection skill learning curves for three methods used by interpreters of high, medium, and low proficiency

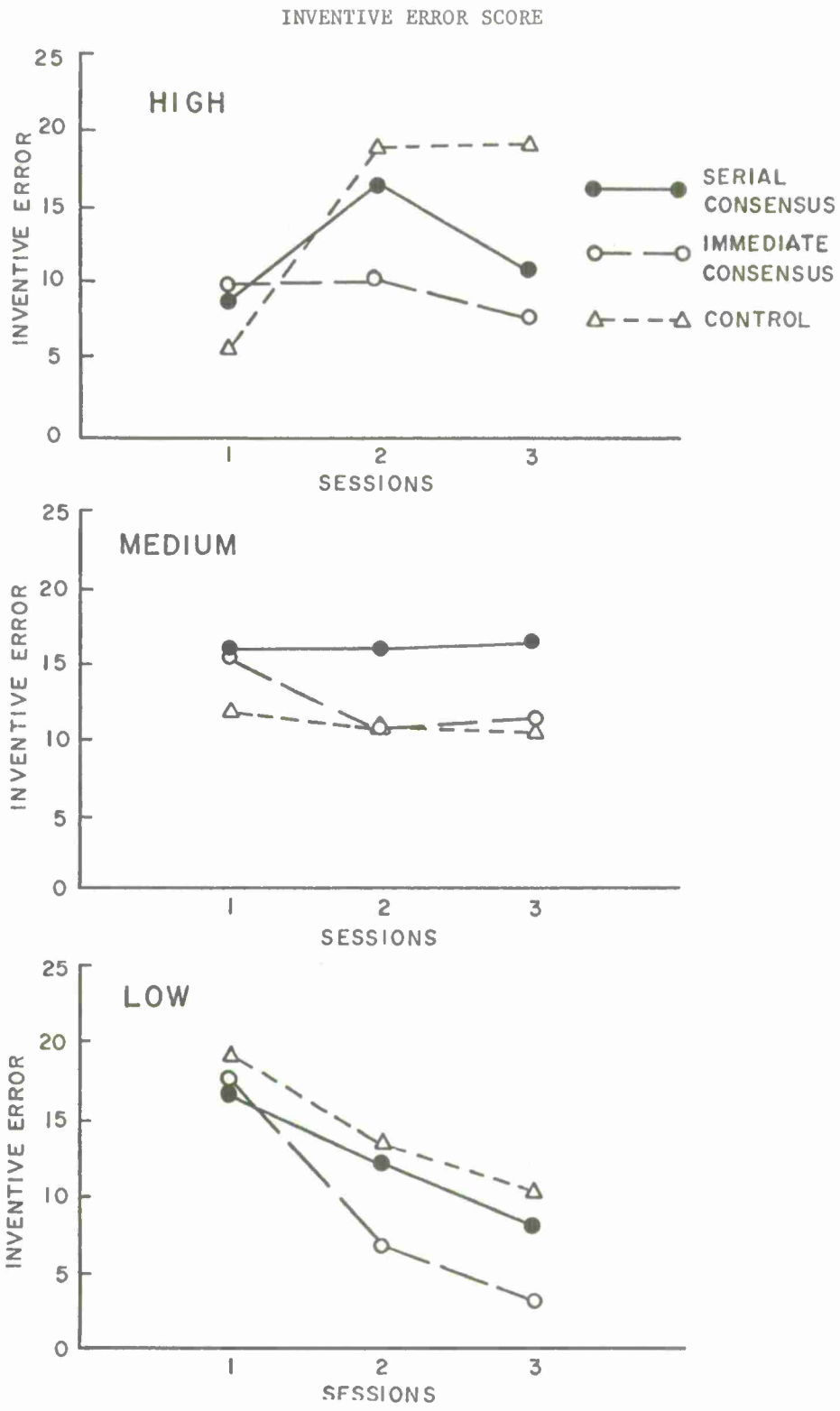

Figure 3. Inventive error learning curves for three methods used by interpreters of high, medium, and low proficiency.

Table 19

P-VALUES FOR THE NEWMAN KEULS-TEST FOR THE THREE SESSIONS SEPARATELY FOR ALL SCORES--EXPERIMENT IV

Score	Comparison	Session 1	Session 2	Session 3
Identification	IC vs Control	ns	.05	.05
	SC vs Control	ns	.05	.05
	SC vs IC	ns	ns	ns
Correct Detection	IC vs Control	ns	ns	.05
	SC vs Control	ns	.01	.05
	SC vs IC	ns	.05	ns
False Target Detection	IC vs Control	ns	ns	ns
	SC vs Control	ns	ns	ns
	SC vs IC	ns	ns	ns

APPENDIXES

APPENDIX A TARGET LIST

T TRACKED VEHICLES

- TT Tanks
- TS SP (Guns, Howitzers, Mortars, Anti-aircraft)
- TA APC's
- TB Armored Bridge Launchers
- TR Recovery Vehicles
- TP Prime Mover/Tractor

A ARTILLERY

- AT Towed Howitzers
- AM Mortar
- AA Anti-aircraft
- AK Anti-tank

M MISSILES

- MS Surface-to-Surface Missile
- ML Missile Launcher/Transporter
- MT Missile Transporter
- MA SAM
- MM SAM Launcher/Transporter

W WHEELED AND CONSTRUCTION VEHICLES

- WL Light Cargo Trucks, 1/4 Ton, 3/4 Ton, Ambulance
- WH Heavy Cargo Trucks, 2-1/2 Ton, 5 Ton, 10 Ton
- WK Tank Trucks (Water, Fuel)
- WW Wrecker Trucks
- WT Truck Tractor (List Separate from Trailer)
- WV Van Trucks (Generator, Shop, Communication, Radar)
- WD Dump Truck
- WC Construction Vehicles (Bulldozers, Cranes, Shovels, Scoops, etc.)

L TRAILERS (ANNOTATE SEPARATELY FROM TRUCKS EVEN IF ATTACHED)

- LL Light Cargo, 1/4 Ton, 3/4 Ton
- LH Heavy Cargo, 1-1/2 Ton
- LS Small Special Purpose (Ammo, Generator, Water, Fuel)
- LR Large Special Purpose (Lo Boy, Tank Transporter, Van, Tanker)
- LE House Trailers (Military)

C CANVAS SHELTER

- CS Small Personnel Tents (Pup, Wall)
- CM Medium Special Purpose Tents (CP, Hex, Kitchen)
- CL Large Tents (GP, Maintenance, Hospital)
- CC Miscellaneous (Latrine, Canvas Shelter, Canvas Water Tank, Canvas Covered Supplies, Canvas Covered Garbage Pits, Flys)

APPENDIX B

DATA SHEET[a]

NAME ________ MAN NUMBER ________ STEREO PAIR NUMBER ________ SCALE ________ START TIME ________ STOP TIME ________

1	2	3	4	5	6	7	8
Annot Number	Right Side	Initial Conf.	1st TM Conf.	2nd TM Conf.	Revised Conf.	Final Conf.	Team Score

[a] The data sheet has been reduced in size for reproduction. The original data sheets were printed on 10" x 14" CLEARPRINT. This is a transparent paper which can be placed on the light table and the exact position of each target marked with a lead pencil. The headings shown on the right were varied to meet the needs of each experiment.

DISTRIBUTION

BESRL Distribution List

*1 OASD (M&RA)
1 HQDA (DACS-ZC)
*1 HQDA (DAPE-MPC)
1 HQDA (DAPE-CPS)
*1 HQDA (DAPE-DW)
2 HQDA (DARD-ARZ-B)
1 HQDA (DARD-ARB)
*1 HQDA (DARD-ARP)
*2 HQDA (DARD-OT)
1 HQDA (DARD-IST-P)
5 HQDA (DAFD-DC)
1 HQDA (DAFD-ZBD)
1 HQDA (DAFD-MF)
1 HQDA (DAFD-CN)
1 HQDA (DAMI-DCS)
5 HQDA (DAPO-PM)
1 HQDA (DAIO-ZC)
1 HQDA (DAAG-PSS-A)
1 HQDA (DAAG-ASL)
1 HQDA (DAPM-MSB)
*1 HQDA (DAPM-PL)
*1 HQDA (SGRD-MDB)
3 HQDA (EPEEC)
4 HQDA (RDSE)
1 HQDA (RDRE)
1 HQDA (RDRD)
*1 HQ USAMC, ATTN: AMCRD-TE
2 HQ USCONARC, ATTN: GI-ATPER
10 HQ USCONARC, ATTN: ATIT-STM
2 HQ First US Army
2 HQ Third US Army
2 HQ Fifth US Army
2 HQ Sixth US Army
1 HQ Sixth US Army, ATTN: Library
*6 HQ USACDC, ATTN: DCS/Opns OAD
1 HQ USACDC, ATTN: Library
1 HQ USASA, ATTN: IARD-T
*1 HQ USA Computer Sys Cmd, ATTN: CSCS-TMT-SD
*1 HQ USA Recruiting Cmd, ATTN: USARCO
1 Dir, Project MASSTER, Ft Hood
*1 USA Electronics Cmd, ATTN: AMSEL-CT-A
*1 USA Electronics Cmd, ATTN: AMSEL-CB-AS
*1 USA Electronics Cmd, ATTN: AMSEL-PA-AH
1 USACDC Exper Cmd, Ft Ord, ATTN: Plans Officer
*1 USACDC Exper Cmd, Ft Ord, ATTN: Library
*1 USACDC, Cbt Sys Gp, ATTN: CSYGC-S
*1 USACDC Institute of Land Combat, ATTN: Library
1 USACDC Armor Agency
1 USACDC Artillery Agency
1 USACDC Aviation Agency
1 USACDC Infantry Agency, ATTN: CAGIN-OA
1 USACDC Infantry Agency, ATTN: Central Files
1 USACDC SW Agency
1 USACDC Pers & Adm Svcs Agency
*1 USACDC Liaison Ofc, USAF, Tact Air Warfare Cen, Eglin AFB
1 USA Adv Mat Concepts Agency, ATTN: AMXAM-TS
1 WRAIR, ATTN: Neuropsychiatry Div.
*1 WRAIR, ATTN: Dept of Exper Psycho-physiology
*1 USA Med Res Lab, Ft Knox, ATTN: Exper Psychol
*1 USA Natick Labs, ATTN: Library
1 USA Natick Labs, ATTN: Psychol Labs
1 USA Natick Labs, ATTN: Assoc Dir, Beh Sci, Pion Div
*1 USA Mobility Eq. R&D Cen, ATTN: HFE
*1 USA Mobility Eq. R&D Cen, ATTN: Library
*1 USA Electronic Prvg Grd, ATTN: STEEP-SU-I
*1 USA Electronic Prvg Grd, ATTN: STEEP-T-BI
2 USA Aberdeen Prvg Grd, ATTN: Dir, HEL
*1 USA Aberdeen Prvg Grd, ATTN: HEL, R. A. Monty
*1 USA Edgewood Arsenal, ATTN: Hum Factors Gp, Biomed Lab
1 USA Tropic Test Cen, Ft Clayton, ATTN: STETC-PD-M
2 USA Arctic Test Cen, Alaska
*1 USA Gen Eq. Test Activ., Ft Lee, ATTN: HF Div, MED
*1 USA Avn Mat Lab, ATTN: Library
*1 EWL, Ft Holabird, ATTN: Intel Mat. Dev. Off.
*2 USA Garrison, Ft Huachuca, ATTN: Tech Ref Off
1 USA C&GSC, ATTN: Educ Adv
6 USA C&GSC, ATTN: Library
1 USA C&GSC, ATTN: Beh Sci Rep, Dir Grad Stud & Res
1 USA C&GSC, ATTN: DOI
*2 USMA, Mil Psych & Ldrshp
*1 USMA, Dir of Institutional Res
1 USMA, Library
1 USA Ordn Cen & Sch, Aberdeen PG, ATTN: DOI
1 USA Ordn Cen & Sch, Aberdeen PG, ATTN: ATSOR-1R
2 USA Engr Sch, Ft Belvoir, ATTN: DOI
*2 USA Mgt Sch, Ft Belvoir
1 USA Inf Sch, Ft Benning, ATTN: DOI
*1 USA Inf Sch, Ft Benning, ATTN: AJIGP
*1 USA Inf Sch, Ft Benning, ATTN: AJIIS-H
1 USA Inf Sch, Ft Benning, ATTN: Library
1 USA Inf Bd, Ft Benning, ATTN: Secy ATBD
1 USA Air Def Sch, Ft Bliss, ATTN: DOI
*2 USA Air Def Sch, Ft Bliss, ATTN: Library
1 USA Air Def Sch, Ft Bliss, ATTN: ATSAD-DL-C
*1 USA Air Def Bd, Ft Bliss
1 USA Inst Mil Asst, Ft Bragg, ATTN: DOI
1 USA Inst Mil Asst, Ft Bragg, ATTN: ATSMA-ID
1 USASA Tng Cen & Sch, Ft Devens
2 USA Transport Sch, Ft Eustis, ATTN: DOI
2 USA MP Sch, Ft Gordon, ATTN: DOI

*Does not receive reports for all research areas.

1 USA SE Sig Sch, Ft Gordon, ATTN: DOI
1 USA SE Sig Sch, Ft Gordon, ATTN: Ed Advsr
2 USA Chaplains Sch, Ft Hamilton, ATTN: DOI
2 USA AG Sch, Ft Benjamin Harrison, ATTN: DOI
2 USA Fin Corps Sch, Ft Benjamin Harrison, ATTN: DOI
*1 USA Med Fld Svc Sch, Ft Sam Houston, ATTN: Stimson Library
2 USA Cbt Survl Sch Tng Cen, Ft Huachuca, ATTN: DOI
2 USA Int Cen & Sch, Ft Huachuca, ATTN: DOI
*1 USA Int Cen & Sch, Ft Huachuca, ATTN: AHBH-S-CB-(ASRC)
*1 USA Int Cen & Sch, Ft Huachuca, ATTN: Library
1 USA Armor Sch, Ft Knox, ATTN: DOI
*1 USA Armor Sch, Ft Knox, ATTN: AHBAAS-DMP
2 USA QM Sch, Ft Lee
2 USA Chem Corps Sch, Ft McClellan
2 USWAC Sch, Ft McClellan, ATTN: DOI
1 USA Sig Cen & Sch, Ft Monmouth, ATTN: DOI
1 USA Sig Cen & Sch, Ft Monmouth, ATTN: Educ Advsr
1 USA Msl & Mun Cen & Sch, Redstone Arsnl, ATTN: DOI
*1 USA Msl & Mun Cen & Sch, Redstone Arsnl, ATTN: Lib
2 USA Avn Sch, Ft Rucker, ATTN: DOI
1 USA Avn Sch, Ft Rucker, ATTN: Library
*1 USA Avn Test Bd, Ft Rucker, ATTN: STEBG-TP-V
*1 USA Fld Arty Sch, Ft Sill, ATTN: DOI (Ed Svcs Div)
2 USA Prim Hel Cen/Sch, Ft Wolters, ATTN: DOI
2 ONR, Code 458
3 US Navy, BuPers (Pers A3)
*1 US Navy, BuM&S, Code 513
*1 US Navy, BuM&S, Code 713
1 Nav Pers Tng Rsch Lab, San Diego, ATTN: Library
1 Nav Pers Tng Rsch Lab, San Diego, ATTN: Stat Dept
*3 Nav Pers Tng Rsch Lab, San Diego, ATTN: Pers Meas Rsch Dept
2 Nav Pers R&D Lab, Washington
*1 Nav Med NA Rsch Unit, San Diego
15 Nav Rsch Lab, Library, Code 2029
1 Nav Rsch Lab, Code 2027
*1 Nav Tng Dev Cen, Orlando, Code 55
*1 Nav Aerospace Med Inst, Pensacola, Code 31
1 Nav Postgraduate Sch, Monterey, ATTN: Library
*2 HQ USMC, ATTN: ADIB
*1 HQ USAF (AF/DPXYA)
*1 HQ USAF (INYMA)
1 HQ AFSC/DLSL
*1 HQ AFLS DPCB
*1 SMAMA(DPCB)
1 USAF Air University AUL/LSA
3 USAF School of Aerospace Med-Aeromedical Lib (SCL-4)
3 US Air Force Academy, PDO 4013A
*1 US Air Force Academy, DMPL
1 AFHRL (DOJZ), Brooks AFB
*1 AFHRL (FT), Williams AFB
1 AFHRL/MD, Alexandria, Va.

*3 AFHRL (PD), Lackland AFB
*1 CINCPACAF (IGSAG)
2 US Armed Forces Institute
*1 American Embassy NATO International Staff
*1 US Dept of Agriculture, Pers Rsch Staff
*1 US Dept of Labor, Info Cen for Mpr Rsch
2 US Coast Guard Hq, C/Psychol Rsch
*1 US Coast Guard Acad., Director of Admissions
*1 US Coast Guard Tng Cen., Governors Island, Educ Advis.
*1 National Security Agency, ATTN: M13
*1 Central Intelligence Agency, ATTN: OCR/ADD/Publications
*1 Nat Cen for Child Health and Human Dev
1 NIMH OC NCMHI
*1 FASEB, Life Sciences Rsch Ofc
*1 BLM, Denver
2 HumRRO, Alex., ATTN Library
2 HumRRO Div 2, Ft Knox
1 HumRRO Div 3, Presidio of Monterey
1 HumRRO Div 4, Ft Benning
*1 HumRRO Div 5, Ft Bliss
1 HumRRO Div 6, Ft Rucker
1 Inst. for Def Anal., Arlington
1 Rsch Anal Corp, McLean, ATTN: Library
1 RAND Corp, Santa Monica, ATTN: Library 98D
*1 Cen for Nav Anal, Arlington, ATTN: Document Center
*1 AIR CRESS, Kensington, ATTN: Info Sys Br
1 Hum Engr Info & Anal Svc, Tufts Univ
*1 Biological Abstracts
1 Univ Illinois, Aviation Rsch Lab
*1 Human Factors Rsch Inc, Goleta
1 Litton Sci Sup Lab, Ft Ord, Behavioral Sciences
140 Library of Congress, Unit Doc Exped Proj
6 Library of Congress, Exchange & Gift Div
12 Defense Documentation Center
*4 Dir Psych, Army Hq, Russell Ofcs, Canberra
1 Scientific Advsr, Mil Bd, Army Hq, Russell Ofcs, Canberra
1 Mil and Air Attacha, Austrian Embassy
*1 Centre de Recherche Des Facteurs, Humaine de la Defense Nationale, Brussels
4 British Defense Staffs Washington, ATTN: Scientific Officer
2 Canadian Joint Staff Washington
3 C/Canadian Defense Rsch Staff, C/CRDS(W)
1 C/Air Staff, Royal Canadian AF, ATTN: Pers Std Anal Br
*1 Def & Civ Inst of Environ Med, Canada
*1 Militaerpsykologisk Tjeneste, Copenhagen
*1 Military Attache, French Embassy, ATTN: Doc Sec
*1 Medecin Chef, C.E.R.P.A.-Arsenal, Toulon/Naval France
1 Prin Scientific Off, Appl Hum Engr Rsch Div, Ministry of Defense, New Delhi
1 Pers Rsch Ofc Library, AKA Irgun, Israal Defense Forces
*1 Ministerie van Defensie, Oper Officer Personeal KL, Afd Sociaal Psychologische Zaken, The Hague, Netherlands

*Does not receive reports for all research areas.

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) System Development Corporation, Falls Church, Virginia and Behavior and Systems Research Laboratory, Arlington, Virginia.	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

TRAINING INDIVIDUAL IMAGE INTERPRETERS USING TEAM CONSENSUS FEEDBACK

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

5. AUTHOR(S) (First name, middle initial, last name)

John T. Cockrell (SDC) and Robert Sadacca (BESRL)

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
June 1971	54	6

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
DAHC 19-67-C-0040 b. PROJECT NO. DA R&D PJ No. 2Q662704A721 c. Surveillance Systems d. e-21	Technical Research Report 1171 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Asst Chief of Staff for Intelligence, Wash., D. C. Office, Chief of Research and Development, DA, Wash., D. C.

13. ABSTRACT

In a series of experiments conducted over the past several years, the ADVANCED SURVEILLANCE SYSTEMS Work Unit of the Behavior and Systems Research Laboratory (BESRL) has sought to develop and test the team consensus feedback method as a technique for maintaining and enhancing the proficiency of individual image interpreters. In the present research, conducted jointly by the BESRL unit and personnel of the System Development Corporation, four experiments were undertaken to explore various team consensus feedback methods with varying conditions of team size and composition. Specifically, the experiments were directed to developing team practice methods which would lead to the greatest performance gains by individual interpreters in an advanced interpretation system.

The first experiment was designed to obtain a general assessment of the usefulness of the consensus feedback process in both target detection and target identification. The second experiment sought to determine if team consensus feedback would be effective in enhancing target detection performance only. Experiment III evaluated modified team consensus feedback techniques designed to minimize feedback delays. Experiment IV was concerned mainly with the nature of learning curves using team consensus feedback techniques and with evaluation of the most effective procedures identified in the first three experiments.

Team consensus feedback was demonstrated to be an effective method of maintaining and improving image interpreter proficiency in an operational image interpretation facility. Greatest performance gains were made in target identification. Results

DD FORM 1 NOV 68 1473 REPLACES DD FORM 1473, 1 JAN 64, WHICH IS OBSOLETE FOR ARMY USE.

Unclassified

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
*Surveillance systems research						
Image interpreter skills						
*Feedback techniques						
Target detection						
*Consensus						
Target identification						
Experimentation						
*Image interpreter proficiency						
*Team consensus feedback method						
Team discussion						
Team size						
Team composition						
Operational image interpretation facility						
Tactical imagery						
Intelligence information						
*Military psychology						
Laboratory facilities						

DD Form 1473

13. ABSTRACT continued

showed that interpreters who were initially low in proficiency achieved the most significant gains in performance. Low proficiency interpreters learned from, or benefited through their collaboration with, higher proficiency interpreters. Neither team discussion nor team size significantly affected interpreter performance; however, these variables did combine to produce an overall effect. The ideal procedure seems to be one using three-man teams, heterogeneous in terms of initial proficiency, in which individuals first perform initial interpretation on an individual basis and then discuss their identifications freely in deciding on the team report.